3-4-2015

Linking Electronic Health Records to Better Understand Breast Cancer Patient Pathways Within and Between Two Health Systems

Caroline A. Thompson
Palo Alto Medical Foundation Research Institute, thompsonc@pamfri.org

Allison W. Kurian
Stanford University School of Medicine, akurian@stanford.edu

Harold S. Luft
Palo Alto Medical Foundation Research Institute, lufth@pamfri.org

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Linking Electronic Health Records to Better Understand Breast Cancer Patient Pathways Within and Between Two Health Systems

Abstract

Introduction: In a fragmented health care system, research can be challenging when one seeks to follow cancer patients as they seek care which can continue for months or years and may reflect many physician and patient decisions. Claims data track patients, but lack clinical detail. Linking routine electronic health record (EHR) data with clinical registry data allows one to gain a more complete picture of the patient journey through a cancer care episode. However, valid analytical approaches to examining care trajectories must be longitudinal and account for the dynamic nature of what is "seen" in the EHR.

Methods: The Oncoshare database combines clinical detail from the California Cancer Registry and EHR data from two large health care organizations in the same catchment area—a multisite community practice and an academic medical center—for all women treated in either organization for breast cancer from 2000 to 2012. We classified EHR encounters data according to typical periods of the cancer care episode (screening, diagnosis, treatment) and posttreatment surveillance, as well as by facility used to better characterize patterns of care for patients seen at both organizations.

Findings: We identified a "treated" cohort consisting of women receiving interventions for their initial cancer diagnosis, and classified their encounters over time across multiple dimensions (type of care, provider of care, and timing of care with respect to their cancer diagnosis). Forty-three percent of the patients were treated at the academic center only, 42 percent at the community center only, and 16 percent of the patients obtained care at both health care organizations. Compared to women seen at only one organization, the last group had similar-length initial care episodes, but more frequently had multiple episodes and longer observation periods.

Discussion: Linking EHR data from neighboring systems can enhance our information on care trajectories, but careful consideration of the complexity of the treatment process and data generating mechanisms is necessary to make valid inferences.

Conclusion/Next Steps: If analyzed as a timeline, and with careful characterization of diagnostic tests, surgical interventions, and type and frequency of physician encounters, the pathways taken by women through their breast cancer episode may lead to better understanding of patient decisions.

Acknowledgements

This submission is based on work presented at the 2014 EDM Forum Symposium. C.A. Thompson was supported by a private grant from the Richard and Susan Levy Foundation.

Keywords

2014 EDM Forum Symposium, Methods, Data Use and Quality

Disciplines

Health Services Research

Linking Electronic Health Records to Better Understand Breast Cancer Patient Pathways Within and Between Two Health Systems

Caroline A. Thompson, PhD, MPH;[i] Allison W. Kurian, MD, MSc;[ii] Harold S. Luft, PhD[i]

Abstract

Introduction: In a fragmented health care system, research can be challenging when one seeks to follow cancer patients as they seek care which can continue for months or years and may reflect many physician and patient decisions. Claims data track patients, but lack clinical detail. Linking routine electronic health record (EHR) data with clinical registry data allows one to gain a more complete picture of the patient journey through a cancer care episode. However, valid analytical approaches to examining care trajectories must be longitudinal and account for the dynamic nature of what is "seen" in the EHR.

Methods: The Oncoshare database combines clinical detail from the California Cancer Registry and EHR data from two large health care organizations in the same catchment area—a multisite community practice and an academic medical center—for all women treated in either organization for breast cancer from 2000 to 2012. We classified EHR encounters data according to typical periods of the cancer care episode (screening, diagnosis, treatment) and posttreatment surveillance, as well as by facility used to better characterize patterns of care for patients seen at both organizations.

Findings: We identified a "treated" cohort consisting of women receiving interventions for their initial cancer diagnosis, and classified their encounters over time across multiple dimensions (type of care, provider of care, and timing of care with respect to their cancer diagnosis). Forty-three percent of the patients were treated at the academic center only, 42 percent at the community center only, and 16 percent of the patients obtained care at both health care organizations. Compared to women seen at only one organization, the last group had similar-length initial care episodes, but more frequently had multiple episodes and longer observation periods.

Discussion: Linking EHR data from neighboring systems can enhance our information on care trajectories, but careful consideration of the complexity of the treatment process and data generating mechanisms is necessary to make valid inferences.

Conclusion/Next Steps: If analyzed as a timeline, and with careful characterization of diagnostic tests, surgical interventions, and type and frequency of physician encounters, the pathways taken by women through their breast cancer episode may lead to better understanding of patient decisions.

Introduction

The rapid wide-scale adoption of the electronic health record (EHR), encouraged by meaningful use initiatives,[1] offers great promise for evidence-based medicine to use routinely collected data on patients of all types, rather than solely relying on randomized controlled trials (RCTs) of selected patients.[2,3] EHR data for cancer research are especially promising because most cancer care is delivered in health care systems, and RCTs do not represent the entire cancer patient population, but only those who are eligible and enroll—about 3 percent of United States cancer patients.[4] EHR data have important limitations for many types of research, however, due to systematic errors arising from wide variation in data reliability. Research using such data sources requires rigorous attention to study design.[5–9] A particular problem with EHR data are instances of nonrandom, completely missing data due to provider use variations or patient migration. Most populations captured in EHR systems are highly dynamic with frequent "in and out" migration based on patient choice, employment, insurance, and geography, and this may blind researchers to some types of care. Gaps in care records and poorly defined source populations can not only lead to difficulties for inference, but may also pose fundamental challenges in identifying (and comparing) appropriate study and target populations.[10] The impact of these problems is heightened when data "missingness" may be related to the focus of study, such as a desire to seek treatment at a different facility because of the disease severity or variation in available treatments.

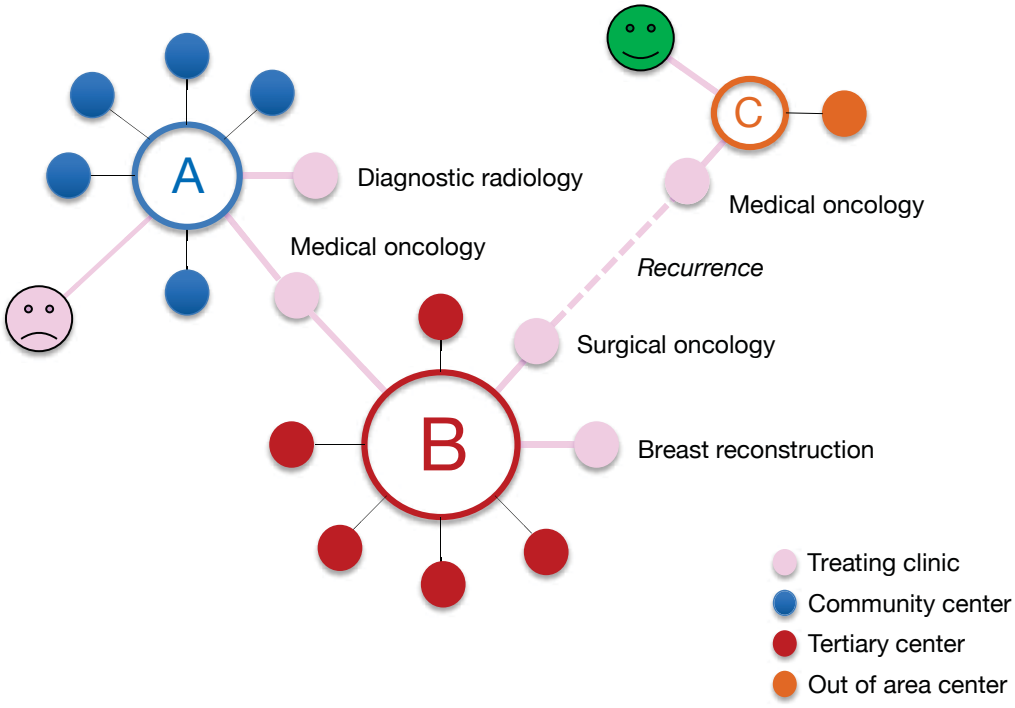
[i]Palo Alto Medical Foundation Research Institute, [ii]Stanford University School of Medicine

Data linkage has been seen as solution to filling the gaps when using incomplete data from a single source.[11–14] Linkage requires finding data on the same persons in multiple EHRs, administrative claims data, or in other population-based data sources such as cancer registries. Cancer registries are particularly useful because they include incidence-based surveillance data and baseline clinical characteristics of disease that are often lacking in, or cannot be easily derived from, EHRs or claims data. Cancer registries, however, generally do not include data on repeated interventions, or cancer recurrences, but these details may be derived from a comprehensive EHR. The SEER-Medicare linkage[15] is an example of linked cancer registry with insurance claims data. An EHR linkage may provide details beyond those in claims data, especially for patient demographics, medical history, and specific medication usage, particularly for information about care *before* the cancer diagnosis. Linking datasets is a necessary, but not sufficient solution to fragmented data; careful consideration of research questions, the appropriateness of available data, and relevance of the population to answer such questions should not be ignored.[11]

Cancers with high rates of survival, e.g., breast cancer, are often characterized by treatment periods, or "episodes" of care, that can continue intermittently for months or years. The beginning of a cancer care episode may not always correspond with the date a patient obtains a definitive diagnosis. The peri-diagnosis phase is often a time of intense resource use that may involve patients seeking care or physician opinions at multiple health care organizations.[16] After the initial episode of treatment—which may involve multiple interventions and cycles, routine posttreatment surveillance may continue indefinitely. The need for repeat courses of therapy or a recurrence of the cancer can result in patients receiving care in multiple facilities. Breast cancer, for example, may be diagnosed in a community-based setting, then treated in a tertiary center, followed by a return to the community setting for surveillance. Thus, the care "pathway" of a breast cancer survivor can be a seen as a "journey" involving multiple providers, procedures, and (possibly) multiple institutions or geographic areas. Each institution and data source, however, may have information on only part of that journey. One key to understanding breast cancer care may be the linkage of EHR from multiple institutions covering long spans of time during this journey (Figure 1).

In this paper we present a case study of our attempts to make sense of longitudinal patterns of breast cancer care after combining EHR data from neighboring academic and community health care systems, augmented by a statewide cancer registry. Considering the extended length of breast cancer treatment and follow-up, we classify the clinical details in the EHR according to the customary cancer care periods (prediagnosis, treatment, and surveillance). We examine the comparability of the patients

Figure 1. Breast Cancer Care in a Fragmented Health Care System



Notes: This figure shows the hypothetical journey or "pathway" of one patient that takes place in three different health care systems (thus three EHRs). In the far left of the figure, the patient, "Mary" presents with breast symptoms to her community health care facility. Mary has a diagnostic workup by her primary care physician and radiology, and she is referred to a medical oncologist, who works at both the community center "A" and at the nearby tertiary academic center "B." The medical oncologist refers Mary to a surgical oncologist and a plastic surgeon, both of whom are part of an academic center. Mary has her surgery, then returns to the community center to be managed by her medical oncologist for her chemotherapy and posttreatment surveillance. Some time passes and Mary moves to a new state (dotted line). She has a recurrence of her breast cancer and is treated by a new medical oncologist at "C"— an out-of-area center.

Ultimately, Mary is cancer free at the end of her journey. Researchers wanting to understand Mary's complete cancer pathway would need clinical details from three EHRs as well as confirmed diagnosis details from a cancer registry. The EHR data from organization A, in particular, might include important prediagnosis information about Mary, such as her frequency of screening, prior concerns, and patterns of noncancer preventive care. Any less detail would limit our understanding of Mary's pathway to shorter time frames, e.g., cancer diagnosis or treatment care periods. If Mary were an Oncoshare patient, we would be able to follow her through her care pathway at "A" and "B," but we would have no information about her recurrence, which was treated at "C."

identified in the two EHR systems with respect to these periods of care. We pay special attention to the subset of patients treated at both organizations due to their apparently higher use of diagnostic and treatment services.

Case Study: Oncoshare

Context

The Oncoshare database combines California Cancer Registry (CCR) data with clinical details from electronic health records (EHR) from two neighboring health care organizations: Palo Alto Medical Foundation (PAMF), a multisite community-based practice; and Stanford University Medical Center (SUMC), an academic medical center. Both are located in the greater Bay Area, in Northern California. SUMC is the closest tertiary center to most PAMF sites—within 1 to 35 miles from each PAMF site that provides cancer care. The Oncoshare patient population includes all women (male breast cancer is not included) diagnosed with or treated for breast cancer at either institution during the period of 2000–2012. The details of how these sources of data were combined have been published in detail elsewhere.[17,18] Briefly, all patients having a physician encounter with an ICD-9 code of breast cancer (174.x) at either institution between 2000 and 2012 were identified and linked to the CCR tumor registry using a probabilistic algorithm based on birthdates, social security numbers, and medical record numbers.[19] Tumor registry data for confirmed cancer patients, including age, race and ethnicity, tumor details, treatment summaries, survival status, and census block data were combined with clinical data extracted from each institution's EHR. The EHR data elements included details of physician encounters, surgical procedures, laboratory and pharmacy orders, medication, and radiotherapy administration records. All personal identifying information was removed before research use of the data, and each patient's clinical encounter dates were offset by a randomly generated factor of -30 to +30 days. The data linkage resulted in a sample of 13,377 women with at least one ICD-9 diagnosis code of breast cancer (174.x) in EHR from either, or both, organization *and* a confirmed diagnosis of breast cancer in the CCR during the same 13-year period.

Using an earlier subset of this cohort (12,109 patients diagnosed between 2000 and 2010), Kurian, et al.[18] identified 1,902 patients (15.7 percent) with breast cancer-related physician encounters at both organizations. That cross-sectional, descriptive analysis did not observe remarkable differences in prognostic factors among these women compared to those treated at only one organization. It did find, however, that use of services such as diagnostic imaging, biopsies, surgeries, radiation therapy, and chemotherapy was significantly greater in women seen at both organizations. This initial analysis highlighted what additional information could be gained from combining data sets. It was not focused on understanding the pathways followed by women in their cancer treatment or on distinguishing those women seeking care in the two settings at the same time from those who were cared for first at location A and then, years later, at location B. We are now beginning to explore differential pathways of care requiring us to further characterize when and how these women sought care at both organizations. Shifting the research focus from characterizing patients seen at a particular organization and the services they received to characterizing the pathways taken by patients as they seek care changes how the data will be used and the preparation steps required.

Methods

Analytical Cohort

For this "pathways" project, we sought an analytical cohort based on evidence that a patient had received care at one or more of our organizations for her initial breast cancer treatment (that is, we exclude women whose first appearance in our data is for treatment of recurrence of breast cancer or surveillance after breast cancer treatment). We included, therefore, only patients with at least one intervention for their initial cancer diagnosis (e.g., chemotherapy, radiotherapy, mastectomy) in the EHR. To reduce the chances any cases were missed, the cohort was initially identified using ICD-9 diagnosis codes in the EHR with the CCR linkage blind to the details of the treatment. A large subset of the patients in the complete sample of 13,377 were confirmed as cancer cases, but were not treated for their initial breast cancer diagnosis at one of the facilities. Thus, for the pathways study we excluded patients who were seen in the EHR data only for screening or posttreatment surveillance mammograms, evaluation and management (EM) visits, or were treated at one of the two institutions under study for recurrent cancer initially treated elsewhere. We drew data from EHR encounters covering the years 2000–2012 (dates and types of services such as EM, imaging, biopsies, surgeries, radiation therapy, chemotherapy, etc.), and diagnosis details (date of diagnosis, diagnosis reporting facility) as linked by the CCR.

Data Classification

To identify *breast cancer-related* care patterns, we distinguished cancer-related procedures from services not likely related to cancer care and divided time into typical phases of the cancer episode (screening, diagnosis, treatment, surveillance). We classified each patient encounter as being provided by the academic center or the community center, and we then assigned an affiliation (academic, community, or both) based on where each patient received her breast cancer care. Classification was accomplished by logic-based (as opposed to data-driven) algorithms defined after careful review of the patient-level data. We relied only on coded fields, ignoring information potentially in free-text or physician notes. Classification and descriptive analyses were performed using SAS version 9.3 (SAS Institute, Cary, N.C.).

Procedures

Over a 13-year period the codes used for cancer and related procedures can vary and coding patterns can differ across organizations. We therefore used Healthcare Cost and Utilizations Project's Clinical Classification Software for services and procedures (HCUP-CCS) to convert ICD-9 and CPT codes to 244 major categories of services and procedures.[20] Using HCUP-CCS allowed us to harmonize the data from two EHRs without having to build a specific list of procedure codes, as would typically be done for an EHR-based study of cancer treatment over a short period of

time at a single site. The resulting coded data included both breast cancer and non-breast cancer-related system "touches". (The term "touches" is broader than "encounter" because it includes entries in the EHR that may or may not be perceived by the patient as a "visit," e.g., a biopsy specimen or radiological image can be interpreted by multiple providers without the patient being present. It can also include indications of electronic messaging between the patient and clinicians.) We further classified as breast cancer-related those procedures that were not clearly breast cancer-specific but had an associated ICD-9 diagnosis code of 174.x. The HCUP categories and their associated ICD-9 and CPT procedure codes for breast cancer-related procedures are provided in Appendix Tables 1A and 2A.

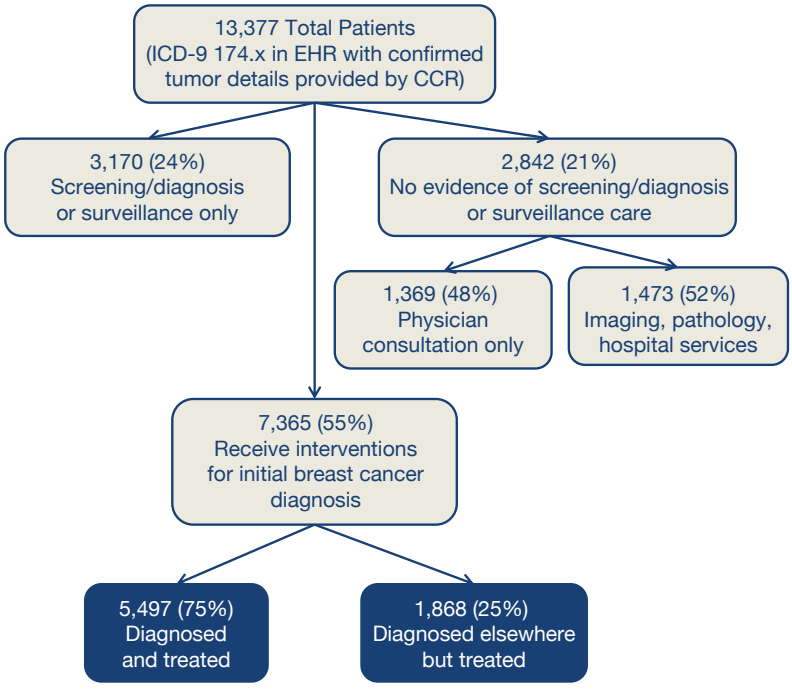
Care Periods

We divided each patient's observation time into care periods using a modified version of the National Cancer Institute cancer control continuum,[21] shown in Figure 2. Each procedure or office visit was classified as a screening, peri-diagnosis, treatment, or posttreatment surveillance encounter based on what was done and when it occurred during the patient's care pathway. The screening period was defined as any time (for which we had data) prior to 90 days before the date of diagnosis (as provided by the CCR). The peri-diagnosis period was defined as the period beginning 90 days prior to this diagnosis date and continuing until the date of the first treatment intervention (the index date). By defining the index date as the first treatment encounter, rather than the CCR-confirmed diagnosis date, our intent was to capture all diagnostic services that may have an impact on treatment choice, including multiple visits to various care providers, some of which may occur *after* the formal diagnosis date. For example, this peri-diagnosis part of the pathway may include visits in response to a suspicious mammogram, then biopsy, perhaps multiple visits dealing with ambiguous results, a definitive diagnosis being made, a search for physicians who might suggest various treatment plans, and then the scheduling of the execution of the first steps in that plan.

Our definition of the treatment period extended from the index date to the point when 180 days had passed without any active treatment. Encounters with cancer specialists that involved imaging, tests, etc., but not active treatment, did not restart the 180-day "clock." If 180 days passed without treatment but routine mammograms continued, the period from the last treatment encounter (less a 30-day buffer around the last treatment encounter) was classified as the posttreatment surveillance period, which continued until another course of therapy was initiated. If another course of therapy was initiated, a new care period was started beginning with a second diagnosis phase at 90 days *before* the first date of the new course of therapy. Treatment (episode 2) continued until we identified another 180-day gap as described above. The diagnosis, treatment, and posttreatment surveillance periods could thus repeat, allowing for multiple cancer-care episodes.

A completed episode was defined as the sequence of diagnosis and treatment, allowing for multiple interventions as long as there was no gap between cancer treatment services exceeding 180 days. The posttreatment surveillance period continued EM visits with a breast specialist) in the EHR data. Each patient's pathway could be unique—some women had multiple treatment episodes, some had very long surveillance periods, other women were lost to follow-up or died during the observation period (deaths were not excluded from this analysis, even if they occurred within an initial treatment period).

Figure 2. Defining the Analytical Cohort



Observation time was characterized in two ways—time for breast-related touches, and total EHR observation time. Breast-related observation time began from the first date of any screening, diagnostic, intervention, or surveillance procedure for routine breast cancer screening, breast symptoms, or the diagnosis of breast cancer, and continued until the most recent breast-related touch (through December 31, 2012). For total EHR observation time (for any procedures, breast cancer-related or other indications), the time was counted from the first encounter in the linked EHR database to the most recent encounter (through December 31, 2012). Thus, if a woman were seen in one of our organizations only for her breast cancer, her total EHR time would be equal to her breast-related observation time. In contrast, if a woman were seen in primary care for many years preceding her breast cancer diagnosis, her total EHR time could be substantially longer than her breast-related observation time. Care period start and stop dates were used to subdivide observation time to determine the average time each patient contributed to each cancer care period, or completed episodes of care.

Organizational Affiliation Classification

Each encounter, care period, and treatment episode was also classified by provider organization. Community breast specialists in some geographic locations use the academic center's inpatient and surgical facilities, and during inpatient care physicians from both organizations can record data in the academic EHR. For patients who appeared in both EHRs, we used an algorithm based on the affiliation of the oncologist or surgeon providing the service. Services from providers such as anesthesiologists, radiologists, pathologists, etc. who were normally involved in a supportive role during a biopsy or major surgery were not considered in this classification. For example, if a patient had a mastectomy by a community surgeon at the academic center the encounter data might also reflect care by academic hospital staff, but the mastectomy would be classified as a community physician-provided service. The academic center also had patients primarily cared for by "private" physicians, i.e., those with admitting privileges who were not employed by either the academic or the community organization. We noted when this was the case.

We aggregated encounters classified in this manner to assign an organizational "affiliation" to each patient for each care period, across all periods within an episode, and across all episodes. We also attempted to identify patients for whom the apparent reason for seeing a physician in "the other organization" was for a second opinion. If all breast cancer-related care in a care period was in one organization, except for a single EM visit to a breast specialist, we classified this patient as visiting both organizations for the purpose of a second opinion only. Final organization affiliation categories were the following: (1) academic, (2) community, (3) both for services, and (4) both for second opinion only (i.e., community patients who visited the academic center for a second opinion or vice versa).

Descriptive Analysis

We provide tables describing the pathways cohort construction, including details regarding the excluded patients. We further describe patients included in the pathways cohort by the organization providing care, care period, and length of follow-up time.

Findings

Creating the analytical cohort for the pathways study (Figure 3; Table 1) resulted in excluding 6,012 patients (45 percent of the source cohort). (Note that because additional data for 2010–2012 became available, the numbers presented here do not match those in Kurian et al. 2014. We applied the original inclusion and exclusion criteria from that paper to the expanded data set, and then narrowed the focus for the purpose of the pathways project.) Eighty-one percent of the excluded women were diagnosed at a facility other than the study facilities; about half of those were diagnosed outside of the four-county area surrounding the study health care organizations. Three thousand one hundred seventy (53 percent) of the excluded women had breast cancer-related E&M visits or procedures in one or both of the EHRs during their screening (e.g., mammography), diagnosis (e.g., biopsy), and surveillance periods (e.g., mammography), but no evidence of treatment during their initial breast cancer episode. Four hundred thirty-three (7 percent) of the excluded women were treated at the academic center by a private physician. The rest of the excluded women (2,842 or 47 percent) had evidence of only one or more E&M visits at the community or academic center, or pathology reports and magnetic resonance imaging services from the academic center (Table 2). The pathways analytical sample thus included 7,365 patients (55 percent of the source cohort) who received interventional treatment for their initial breast cancer diagnosis from one of the two organizations. Of these patients, 5,497 (75 percent) were diagnosed *and* treated in these facilities; 1,868 (25 percent) were initially diagnosed elsewhere, but had evidence of treatment in the EHR. By construction, all 7,365 pathways cohort members were treated during their first cancer episode at one of the two study facilities, 89 percent of the cohort received diagnostic services, 84 percent had evidence of posttreatment surveillance, and 32 percent had evidence of prediagnosis screening (Table 1).

After classification by episode, care period, and provider organization, we identified 3,136 (43 percent) of patients as seen only in the academic setting; 3,059 (42 percent) as seen only in the community settin;, and 1,170 (16 percent) as seen in both settings at any time during our observation period, 2000–2012. Within this pathways cohort, 24 percent of the academic center patients were initially diagnosed outside of the four-county area served by the study facilities, while only 2 percent of the community center patients came from out of area. Compared to academic-center-only patients, community-center-only patients had more EHR evidence of precancer screening services (51 percent versus 10 percent), diagnosis services (97 percent versus 80 percent), and posttreatment surveillance services (92 percent versus 72 percent). Of the women seen in both organizations, 177 (15 percent) appeared to be only seeking a second opinion from the other

organization (of these 39 percent were academic patients seeking opinions from community physicians and 61 percent were community patients seeking opinions from academic physicians; data not shown). Among those obtaining services at both organizations (993 patients), 38 percent had diagnostic procedures at both centers, 36 percent were treated at both centers, and 23 percent had surveillance care at both centers (Table 3). These percentages just happen to come close to 100 percent; 183 women (7 percent) actually received services from both centers in two or more of these categories (not shown).

Table 1. Cohort Construction

	All Patients N (%)		Pathways Cohort N (%)		Excluded N (%)	
All patients	**1,3377**	**(100.0%)**	**7,365**	**(100.0%)**	**6,012**	**(100.0%)**
CCR diagnosis reporting facility:						
Academic	2,688	(19.8%)	2,184	(29.6%)	504	(8.3%)
Community	3,936	(29.4%)	3,313	(44.9%)	623	(10.3%)
Other, inside four-county area	3,498	(26.1%)	1,025	(13.9%)	2,473	(41.1%)
Other, outside four-county area	3,255	(24.3%)	843	(11.4%)	2,412	(40.1%)
Types of services for breast cancer:[1]						
A. Screening period (All)	**2,680**	**(20.0%)**	**2,353**	**(31.9%)**	**319**	**(5.3%)**
Academic	467	(3.4%)	386	(5.2%)	81	(1.3%)
Community	2,150	(16.0%)	1934	(26.2%)	216	(3.5%)
Both	42	(0.3%)	33	(0.4%)	9	(0.1%)
Private	21	(0.1%)	–	–	13	(0.2%)
B. (Peri-) Diagnosis period (All)	**7,964**	**(59.5%)**	**6,534**	**(88.7%)**	**1,381**	**(23.0%)**
Academic	3,363	(25.1%)	2673	(36.2%)	690	(11.4%)
Community	3,901	(29.1%)	3331	(45.2%)	570	(9.4%)
Both	578	(4.3%)	530	(7.1%)	48	(0.7%)
Private	122	(0.9%)	–	–	72	(1.1%)
C. Treatment period (All)	**7,798**	**(58.3%)**	**7,365**	**(100%)**	**433**	**(7.2%)**
Academic	3,497	(26.1%)	3497	(47.4%)	–	–
Community	3,397	(25.3%)	3397	(46.1%)	–	–
Both	471	(3.5%)	471	(6.3%)	–	–
Private	433	(3.2%)	–	–	433	(7.2%)
D. Surveillance period (All)	**8,228**	**(61.5%)**	**6,163**	**(83.7%)**	**2,062**	**(34.3%)**
Academic	3,494	(26.1%)	2504	(33.9%)	990	(16.4%)
Community	4,193	(31.3%)	3201	(43.4%)	992	(16.5%)
Both	530	(3.9%)	458	(6.2%)	72	(1.1%)
Private	11	(0.0%)	–	–	8	(0.1%)
None of the above	–	–	–	–	2,842	(47.3%)

Note: [1]Patients may appear multiple times across A–D

Table 2. Most Common Reasons for Care If "None of the Above" in Table 1 (N=2,842)

	Any Period		Screening		(Peri-) Diagnosis		Treatment		Surveillance	
E&M visit (All)	**1,369**	**(48.2%)**	**24**	**(0.8%)**	**406**	**(14.3%)**	**434**	**(15.3%)**	**657**	**(23.1%)**
Academic	904	(31.8%)	3	(0.1%)	311	(10.9%)	325	(11.4%)	328	(11.5%)
Community	441	(15.5%)	21	(0.7%)	74	(2.6%)	107	93.8%)	327	(11.5%)
Private	24	(0.8%)	–	–	21	(0.7%)	2	(0.1%)	2	(0.1%)
Pathology services (All)	**1,421**	**(50.0%)**	**14**	**(0.5%)**	**1,070**	**(37.6%)**	**182**	**(6.4%)**	**181**	**(6.4%)**
Academic	1,394	(49.0%)	11	(0.4%)	1,063	(37.4%)	176	(6.2%)	175	(6.2%)
Community	27	(1.0%)	3	(0.1%)	7	(0.2%)	16	(0.6%)	6	(0.2%)
Magnetic Resonance Imaging (All)	**331**	**(11.6%)**	**4**	**(0.1%)**	**165**	**(5.8%)**	**96**	**(3.4%)**	**117**	**(4.2%)**
Academic	322	(11.3%)	4	(0.1%)	163	(5.7%)	94	(3.3%)	112	(3.9%)
Community	9	(0.3%)	–	–	2	(0.1%)	2	(0.1%)	5	(0.2%)

Figure 3. Episode and Care Period Definitions for Classification of Cancer-Related Longitudinal Encounters Data

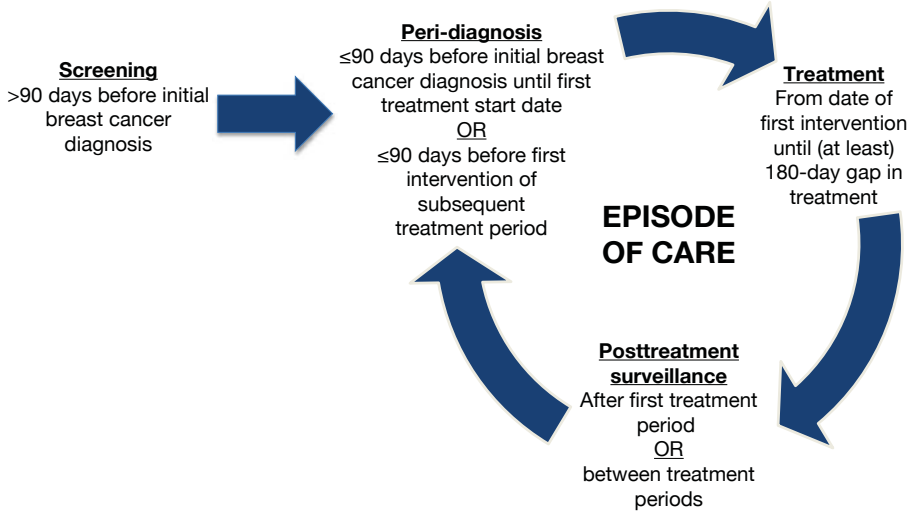


Table 3. Where Services Were Provided, by Care Period, for Any Treatment Episode (N=7,383 Patients)

Where Services Were Provided, by Care Period	Where Services Were Provided by Organization							
	Academic Only		Community Only		Both – Services		Both – Second Opinion Only	
Total	**3,136**	**(100.0%)**	**3,059**	**(100.0%)**	**993**	**(100.0%)**	**177**	**(100.0%)**
CCR diagnosis reporting facility:								
Academic	1,712	(54.5%)	68	(2.2%)	332	(33.4%)	72	(40.6%)
Community	28	(0.8%)	2,642	(86.3%)	567	(57.0%)	76	(42.9%)
Inside four-county area	651	(20.7%)	288	(9.4%)	66	(6.6%)	20	(11.2%)
Outside four-county area	745	(23.7%)	61	(1.9%)	28	(2.8%)	9	(5.0%)
Screening period:								
Academic only	324	(10.3%)	–	–	51	(5.1%)	11	(6.2%)
Community only	–	–	1,577	(51.5%)	321	(32.3%)	36	(20.3%)
Both – Services	–	–	–	–	28	(2.8%)	–	–
Both – Second opinion	–	–	–	–	3	(0.3%)	2	(1.1%)
Neither (other organization)	2,812	(89.6%)	1,482	(48.4%)	590	(59.4%)	128	(72.3%)
(Peri-) Diagnosis period:								
Academic only	2,504	(79.8%)	–	–	122	(12.2%)	47	(26.5%)
Community only	–	–	2,975	(97.2%)	307	(30.9%)	49	(27.6%)
Both – Services	–	–	–	–	384	(38.6%)	–	–
Both – Second opinion	–	–	–	–	85	(8.5%)	61	(34.4%)
Neither (other organization)	632	(20.1%)	84	(2.7%)	95	(9.5%)	20	(11.2%)
Treatment period:								
Academic only	3,136	(100.0%)	–	–	292	(29.4%)	69	(38.9%)
Community only	–	–	3,059	(100.0%)	297	(29.9%)	41	(23.1%)
Both – Services	–	–	–	–	357	(35.9%)	–	–
Both – Second opinion	–	–	–	–	47	(4.7%)	67	(37.8%)
Neither (other organization)	–	–	–	–	–	–	–	–
Surveillance period:								
Academic only	2,264	(72.1%)	–	–	198	(19.9%)	42	(23.7%)
Community only	–	–	2,815	(92.0%)	340	(34.2%)	46	(25.9%)
Both – Services	–	–	–	–	231	(23.2%)	–	–
Both – Second opinion	–	–	–	–	156	(15.7%)	71	(40.1%)
Neither (other organization)	872	(27.8%)	244	(7.9%)	68	(6.8%)	18	(10.1%)

Community patients had longer total EHR observation time and breast cancer-related observation time than academic patients did (Table 4). Treatment periods for women treated in the community site were also longer than for women at the academic site. Women who had services at both organizations had the longest average, overall, and breast-related observation times (we have not yet begun to explore the reasons for these differences.) Six thousand one hundred ninety-three patients (84 percent) of the analytical cohort had only one breast cancer episode; the remaining 1,172 (16 percent) had multiple episodes (Table 4). The first peri-diagnosis and treatment periods for women with more than one episode were not different in length compared to women with only one episode, within organization type. Overall, women with multiple episodes accumulated more observation time in the peri-diagnosis and treatment phases. Further analysis of the 993 women receiving services from both organizations (Table 5) revealed that 262 (26 percent) had more than one episode of care. Of these women, 79 percent were seen at both organizations during the first episode and 21 percent were classified as "both" after completion of their first episode.

Table 4. Episode Length Characteristics (N=7,365 patients)

Observation Time in Months, by Episode and Care Period	Where Services Were Provided by Organization							
	Academic Only		Community Only		Both – Services		Both – Second Opinion Only	
Total observation time, mean (SD):	**(N=3,136)**		**(N=3,059)**		**(N=993)**		**(N=177)**	
Total EHR observation time[1]	63.1	(49.1)	94.9	(40.5)	111.0	(37.7)	99.5	(39.1)
Breast-related observation time[2]	50.7	(44.1)	76.5	(41.7)	86.9	(41.4)	74.1	(41.4)
Observation time by cancer period, mean (SD)								
Patients with 1 episode (N=6193):	**(N=2,570)**		**(N=2,747)**		**(N=731)**		**(N=145)**	
Screening	30.3	(36.2)	30.5	(29.9)	32.6	(33.3)	34.4	(35.1)
Diagnosis	3.0	(9.0)	2.6	(8.7)	4.1	(12.2)	3.4	(10.2)
Treatment	3.7	(3.8)	5.2	(4.7)	5.1	(4.1)	5.6	(4.8)
Surveillance	39.5	(36.9)	45.7	(33.7)	53.2	(38.7)	46.3	(35.6)
Patients with >1 episode (N=1172):	**(N=566)**		**(N=312)**		**(N=262)**		**(N=32)**	
First episode:[3]								
Screening	31.2	(37.1)	19.3	(25.4)	23.0	(27.3)	28.5	(21.8)
Diagnosis	2.9	(8.4)	1.8	(3.1)	3.2	(8.8)	1.3	(1.3)
Treatment	3.4	(3.0)	5.2	(3.6)	4.8	(3.8)	4.9	(3.9)
Surveillance	44.3	(39.5)	48.9	(37.3)	51.0	(42.3)	57.9	(39.7)
Any subsequent episode:[4]								
Diagnosis	1.3	(0.9)	1.4	(0.9)	1.4	(0.9)	1.8	(0.9)
Treatment	2.7	(4.8)	4.4	(6.5)	4.3	(7.9)	4.0	(4.5)
Surveillance	37.4	(34.8)	25.6	(29.8)	37.9	(28.6)	17.3	(18.4)
Cumulative time in all episodes:[5]								
Diagnosis	3.8	(5.9)	3.4	(3.4)	4.8	(9.0)	3.2	(1.4)
Treatment	5.6	(6.3)	9.2	(8.7)	9.4	(11.2)	9.2	(8.4)
Surveillance	50.0	(38.5)	54.0	(35.5)	60.0	(37.4)	62.8	(37.5)

Notes: [1]Total EHR observation time=months from first appearance in EHR (for any reason) to most recent appearance (for any reason), through 2012
[2]Breast-related observation time=months from first breast-related procedure in EHR to most recent breast-related procedure, through 2012
[3]Mean length of each care period for the first episode only for patients with more than one episode
[4]Mean length of any subsequent (after first episode) care period for patients with more than one episode
[5]Mean length of cumulative time in each care period (regardless of episode) for patients with more than one episode

Figure 4. Steps for Care Period Classification for a Patient with Two Breast Cancer Episodes

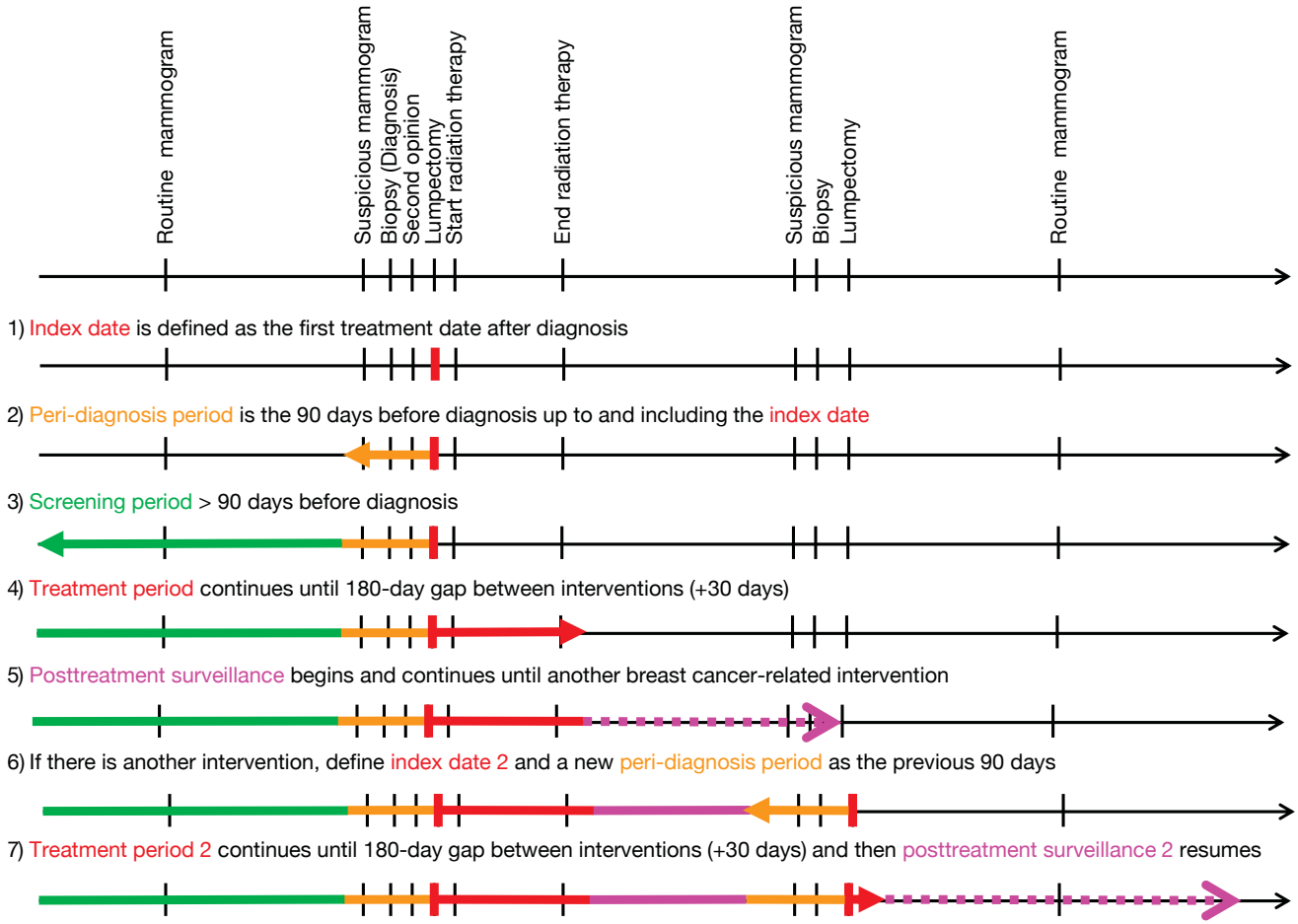


Table 5. Episode Details for Patients Seen at Both Organizations for Services (N=993)

Episode Detail	N (%)
Patients with 1 episode	731 (73.6%)
Patients with >1 episode	262 (26.4%)
Both in first episode	207 (79.0%)
Both after first episode	55 (21.0%)

Discussion

We seek a better understanding of the factors that explain breast cancer care pathways over an observation period of up to 13 years within and between two health care organizations: a tertiary academic center and a neighboring multisite community health care system (those analyses are still underway; the lessons learned in creating the analytic data set for this purpose, however, are of more general value.) To undertake this study we used linked EHR data from both organizations, augmented with diagnosis details from the statewide registry. We defined an analytical cohort based on treatment for an initial breast cancer diagnosis, omitting many other women with breast cancer who were at other points in their pathway. We distinguished breast cancer-related care from other types of care. We determined the point in the care pathway when each breast cancer-related encounter occurred relative to customary cancer care periods (routine screening, peri-diagnosis, treatment, and posttreatment surveillance. We distinguished episodes as completed cycles of care separated by 6-month gaps in care. We classified each patient as receiving care at the academic center, community center, or both; and by care period, episode, and overall, distinguishing whether the care at "both" represented only second opinions.

Although we started with a cohort similar to that used by Kurian et al.,[18] our smaller pathways cohort captured only patients treated in the study facilities for their initial breast cancer diagnosis. The previous analysis identified 52 percent of patients as academic patients, 32 percent as community patients, and 16 percent as patients treated at both organizations. We excluded many patients included in the previous analysis because of onetime consultations, care received only during the diagnosis or posttreatment surveillance period (i.e., patients treated elsewhere), or use of hospital services such as MRI or pathology being the patients' only appearance in the EHR. In our smaller "treated" cohort, we coincidentally found a 16 percent overlap in assigned organizational affiliations, but identified 43 percent as academic and 42 percent community. This change is because of exclusions occurring more

often in the academic setting. By distinguishing reasons for obtaining care, we see that of all the women seen at both organizations, 15 percent were so classified only due to second opinions at the other site; 5 percent, because they had treatment at different sites during *different* episodes; and 80 percent, because they had treatment at both organizations during the same episode.

It is not uncommon to see a substantial reduction in sample from the initial number of cases appearing in the EHR.[10] Indeed, in the previous analysis by Kurian et al., attempts to identify patients in common between the three linked data sources (community EHR, academic EHR, and CCR) reduced sample sizes significantly.[18] To maximize sensitivity, cohorts of EHR patients are often defined by ICD-9 diagnosis codes for procedures or in the problem list. In Oncoshare, this cast a very wide net—including patients at all stages of their cancer care and a large number who received the majority of their care elsewhere. Linkage with CCR tumor confirmation details led to two observations about this ICD-9 code search method of initial cohort definition: (1) over a long follow-up time (e.g., 10+ years) there are likely many more patients with breast cancer diagnosis codes in the EHR (e.g., on the problem list) than with breast cancer treatment details, and (2) in the absence of tumor registry linkage, an EHR-based study of cancer treatment (using treatment details only) may underestimate substantially the period-prevalence of cancer in the system population due to treatment outside the system. Because of the highly curable nature of breast cancer, surveillance time can stretch into decades for many women and includes job changes or relocations that may result in a change of locality or health care provider. Our case study may exaggerate this issue because a large academic medical center such as Stanford is a draw for second opinions and expert care in more complex cases (e.g., recurrences and multiple treatment courses). Indeed, the majority of cases we excluded from our analysis had initial cancer diagnosis at a hospital outside the geographic region commonly served by the study facilities. We expect this issue would be encountered less frequently in a health maintenance organization (HMO) population, where care is provided for all (current) members and data capture may thus be better for as long as the patients remain members. The issue may also be less relevant in studying a cancer with lower survival rates.

In a "researcher-focused ideal world," one would have readily linkable EHR data from all possible providers that could be linked to tumor registry and other data. Aside from the facts that not all organizations have EHRs, that cross-organization patient identifiers do not exist, and that Health Insurance Portability and Accountability Act (HIPAA) privacy concerns would be substantial, few organizations are willing to share what they may consider to be sensitive data with other entities. In the "less than ideal real world" some research questions can nonetheless be addressed with more limited data. Given the Oncoshare partnership between the two organizations, we could have defined our initial study cohort with all cases the CCR would have attributed to each. Beginning with women each organization had identified as potentially having had breast cancer, however, allowed us to see how often such EHR-based measures overidentify cases.

We used the public domain HCUP-CCS algorithm to aggregate procedure codes (ICD-9 and CPT) and ICD-9 diagnosis codes from two different types of organizations. A hierarchical coding system such as CCS (which yields 244 major procedure categories) is useful for harmonizing EHR data, which can be of varying quality and reflect changing coding schemes over time. While there is some loss of clinical detail or granularity in employing such a schema,[22] for our purposes this was a very efficient way to handle the linked data. We were able to classify and retain EHR encounter details for an entire patient pathway, both cancer-related and otherwise.

One-sixth of our patients had multiple treatment episodes—meaning they were initially diagnosed, were treated, and had at least a 6-month gap in treatments that was followed by another diagnostic workup and a second (or subsequent) round of treatment. We saw that, as would be expected, women with multiple episodes accumulated more total observation time in the diagnosis and treatment care periods than did those with only one episode. However, the length of the first episode of women who experienced multiple episodes was not different from that of the episodes of women who had only one episode. There were, however, marked differences between treatment period lengths when we compared academic to community patients, with shorter follow-up times in the former. This was most likely due to academic patients completing courses of radiotherapy or chemotherapy at a community institution (other than the one under study).

We did not begin the cancer treatment episode on the date of diagnosis, but allowed the diagnostic period to extend to the start of the first treatment. Hornbrook et al. observed an increase in health care cost and utilization in the 3–5 month period prior to the month of cancer diagnosis.[16] Part of that period is included in our peri-diagnosis notion, which extends up to 90 days prior to the date of diagnosis. Understanding the factors that are associated with more extensive service use during the diagnostic period, as well as choices during the treatment period was the focus of our pathways study.

Our approach to classification along several dimensions let us retain as much information about a patient's care pathway as possible while allowing flexibility in categorization based on our research questions. Our notion of a treatment episode that is extended by the observation of new interventions within a given window of time is similar to that used in certain commercially available groupers. We chose to use our own logic to be able to control what would, and would not, extend the treatment episode, e.g., surveillance mammograms do not extend the treatment episode.

Conclusion and Next Steps

We believe that linking EHR data offers substantial advantages for the questions we seek to address. For example, a woman's prior experience with screening results, her other conditions, or her long-term relationship with her primary physicians may all

have an impact on the care she received during the diagnostic and treatment phases of her breast cancer care. The EHR data may include information on explicit referrals to another physician, as distinct from a purely patient-driven choice. For such measures EHR data are critical, and more can be learned by linking data from EHRs in neighboring organizations than from just one, but care may be obtained from yet other settings. The state cancer registry offers a unique opportunity to identify such instances, since the registry receives care reports from all facilities in the state, rather than just from the two in this study. Creating a longitudinal data set with the ability to reduce the likelihood of missing data (i.e., omitting women without treatment in either setting) is critical for our purposes. Restructuring the data to understand pathways of care allows us to examine such questions in an efficient and valid manner. The proposed approach and methods discussed have provided opportunities, not just to answer our questions, but to have potential use for others.

Acknowledgements

This submission is based on work presented at the 2014 EDM Forum Symposium. C.A. Thompson was supported by a private grant from the Richard and Susan Levy Foundation.

References

1. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. The New England journal of medicine. 2010;363(6):501-4.
2. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association : JAMIA. 2007;14(1):1-9.
3. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Annals of internal medicine. 2009;151(5):359-60.
4. Warner J, Hochberg E. Where is the EHR in oncology? Journal of the National Comprehensive Cancer Network : JNCCN. 2012;10(5):584-8.
5. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care. 2013;51(8 Suppl 3):S30-7.
6. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. Medical care. 2013;51(8 Suppl 3):S80-6.
7. Stuart EA, DuGoff E, Abrams M, Salkever D, Steinwachs D. Estimating causal effects in observational studies using Electronic Health Data: Challenges and (some) solutions. Egems. 2013;1(3).
8. Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Medical care. 2010;48(6 Suppl):S114-20.
9. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association : JAMIA. 2013;20(1):144-51.
10. Overhage JM, Overhage LM. Sensible use of observational clinical data. Statistical methods in medical research. 2013;22(1):7-13.
11. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health services research. 2010;45(5 Pt 2):1468-88.
12. Manion FJ, Harris MR, Buyuktur AG, Clark PM, An LC, Hanauer DA. Leveraging EHR data for outcomes and comparative effectiveness research in oncology. Current oncology reports. 2012;14(6):494-501.
13. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(34):4243-8.
14. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, et al. Building a research consortium of large health systems: the Cancer Research Network. Journal of the National Cancer Institute Monographs. 2005(35):3-11.
15. National Cancer Institute: SEER-Medicare linked database. Available from: http://appliedresearch.cancer.gov/seermedicare/.
16. Hornbrook MC, Fishman PA, Ritzwoller DP, Elston-Lafata J, O'Keeffe-Rosetti MC, Salloum RG. When does an episode of care for cancer begin? Medical care. 2013;51(4):324-9.
17. Weber SC, Seto T, Olson C, Kenkare P, Kurian AW, Das AK. Oncoshare : lessons learned from building an integrated multi-institutional database for comparative effectiveness research. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:970-8.
18. Kurian AW, Mitani A, Desai M, Yu PP, Seto T, Weber SC, et al. Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. Cancer. 2014;120(1):103-11.
19. Weber SC, Lowe H, Das A, Ferris T. A simple heuristic for blindfolded record linkage. Journal of the American Medical Informatics Association : JAMIA. 2012;19(e1):e157-61.
20. Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS): U.S. Agency for Healthcare Research and Quality; 2013 [cited 2014 8/9/2014].
21. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Statistics in medicine. 2000;19(3):335-51.
22. Le HV, Poole C, Brookhart MA, Schoenbach VJ, Beach KJ, Layton JB, et al. Effects of aggregation of drug and diagnostic codes on the performance of the high-dimensional propensity score algorithm: an empirical example. BMC medical research methodology. 2013;13:142.

Appendix

Table 1A. Diagnosis Codes Used for Calculation of in Breast-Related Observation Time

Diagnosis Category	CCS	ICD-9 Codes
Nonmalignant breast conditions	167	610.0, 610.1, 610.2, 610.3, 610.4, 610.8, 610.9, 611.0, 611.1, 611.2, 611.3, 611.4, 611.5, 611.6, 611.71, 611.72, 611.79, 611.8, 611.81, 611.82, 611.83, 611.89, 611.9, 612.0, 612.1, 793.8, 793.80, 793.81, 793.82, 793.89
Cancer of breast[1]	24	174.0, 174.1, 174.2, 174.3, 174.4, 174.5, 174.6, 174.8, 174.9, 233.0, V103
Secondary malignancies	42	196.0, 196.1, 196.2, 196.3, 196.5, 196.6, 196.8, 196.9, 197.0, 197.1, 197.2, 197.3, 197.4, 197.5, 197.6, 197.7, 197.8, 198.0, 198.1, 198.2, 198.3, 198.4, 198.5, 198.6, 198.7, 198.81, 198.82, 198.89, 209.71, 209.72, 209.73, 209.74, 511.81, 789.51
Malignant neoplasm without specification of site	43	199.0, 199.1, 199.2, 209.20, 209.29, 209.30, 209.70, 209.75, 209.79

Note: [1]Male breast cancer (ICD-9 175.0 and 175.9) are also included in CCS category 24 but they are not applicable in this study because the cohort restricted to female patients.

Table 2A. Procedure Codes Used for Classification of Index Date and Organization Affiliation

Procedure Category	CCS	ICD-9 Procedure Code(s)	CPT Code(s)
Breast cancer treatment:			
Lumpectomy; quadrantectomy of breast[1]	166	85.20, 85.21, 85.22, 85.23	0301T, 19120-19126, 19160-19162, 19297, 19301, 19302
Mastectomy[1]	167	85.41, 85.42, 85.43, 85.44, 85.45, 85.46, 85.47, 85.48	19180-19240, 19300, 19303-19307
Radiation therapy[1,2]	211	92.20, 92.21, 92.22, 92.23, 92.24, 92.25, 92.26, 92.27, 92.28, 92.29, 92.41	0007T, 0073T, 0082T, 0083T, 0182T, 0197T, 41019, 50959, 50978, 74235, 76950-76965, 77261-77334, 77338, 77371-77799, 79030-79403, C2683, C2639, C2640-C2643, C2698, C2699, C9714, C9715, C9725, C9728, G0174, G0178, G0242, G0243, G0251, G0256, G0261, G0273, G0274, G0338- G0340, G0458, S8030, S8049
Cancer chemotherapy[1,2]	224	00.10, 00.15, 17.70, 99.25, 99.28	51720, 61517, 96400-94520, 96530-96549, G0355-G0360, G0361, G0362, Q0083-Q0085, S2107
Evaluation and management:			
Other diagnostic procedures[2,3]	227	00.58, 00.59, 00.67, 00.68, 00.69, 89.01, 89.02, 89.03, 89.04, 89.05, 89.06, 89.07, 89.08, 89.09, 89.10, 89.11, 89.12, 89.13, 89.15, 89.16, 89.17, 89.18, 89.19, 89.21, 89.22, 89.23, 89.24, 89.25, 89.26, 89.31, 89.32, 89.33, 89.34, 89.35, 89.36, 89.37, 89.38, 89.39, 89.45, 89.46, 89.47, 89.48, 89.49, 89.50, 89.53, 89.55, 89.56, 89.57, 89.58, 89.59, 89.61, 89.62, 89.63, 89.66, 89.67, 89.68, 89.69, 89.7, 89.8	0074T, 0089T, 0185T, 0188T, 0203T-0204T, 59420-59430, 77336, 77370, 86077-86079, 88325, 90951-90970, 92521-92524, 92548, 94780-94781, 95105, 95782-95811, 95105, 95782-95811, 95828, 96004, 96040, 99026-99050, 99052, 99054-99058, 99090-99091, 99151-99170, 99201-99285, 99291-99435, 99438, 99441-99463, 99468-99480, 99487-99499, 99605-99607, C9801, C9802, G0246, G0250, G0344, G0380, G0382-G0384, G0390, G0398, G0399, G0402, G0406-G0408, G0425-G0427, G0437-G0439, G0445-G0447, G0449, G0452, G0424, G9156, H0034, H1000-H1005, H1010, H1011, H2000, H2010, Q5010, S0260, S0302, S0353, S0354, S0605, S0610, S0612, S0613, S0622, S5190, S9110, S9117, T1015

Table 2A. Procedure Codes Used for Classification of Index Date and Organization Affiliation (Cont'd)

Procedure Category	CCS	ICD-9 Procedure Code(s)	CPT Code(s)
Screening, diagnosis, and/or surveillance:			
Breast biopsy and other diagnostic procedures on breast	165	85.11, 85.12, 85.19	0046T, 0047T, 0060T, 19000, 19001, 19030-19103, 19281-19295
Computerized axial tomography (CT)[2]	177, 178, 179, 180	87.03, 87.41, 88.01, 00.31, 87.71, 88.38	70450-70488, 70496, 71250-71275, 75571-75573, S8093, 72191-72194, 74150-74178, 74261-74263, 0042T, 0144T, 0146T-0150T, 70490-70492, 72125-71233, 73200-73206, 73700-73706, 76070, 76071, 76355-76375, 76380, 76497, 77011-77014, 77078, 77079, G0131, G0132, S8092
Mammography	182	87.36, 87.37	76082-76092, 77053-77057, G0202-G0207, G0236, S8075, S8080
Routine chest X-ray[2]	183	87.44	0174T, 0175T, 71010-71022, 71030, 71035
Diagnostic ultrasound[2]	192, 193, 194, 195, 196, 197	00.21, 88.71, 00.24, 88.72, 88.74, 00.25, 88.75, 88.76, 00.22, 00.23, 00.28, 00.29, 88.73, 88.77, 88.78, 88.79	0126T, 76506-76536, 93875-93893, S9024, 76825-76828, 76930-76932, 92978, 92979, 93303-93352, 93662, C8921-C8930, 76975, 79776, 76938, G0050
Nuclear medicine imaging[2]	207, 208, 209, 210	92.14, 92.15, 92.01, 92.02, 92.03, 92.04, 92.05, 92.09, 92.11, 92.12, 92.13, 92.16, 92.17, 92.18, 92.19	78102-78104, 78300-78399, 78579-78599, 0331T, 0332T, 70015, 72285, 72295, 78000-78099, 78195, 78202, 78206, 78216-78220, 78226, 78227, 78232-78258, 78262-78276, 78280-78282, 78291, 78414-78428, 78451-78454, 78461, 78465, 78468-78483, 78492-78496, 78601, 78606, 78608-78999, G0031, G0034-G0047, G0125, G0126, G0163-G0165, G0210-G0234, G0252, G0253, G0296, G0336, S8004, S8085, 78110-78193, 78199-78201, 78205, 78215, 78223, 78230-78231, 78261, 78278, 78290, 78299, 78445, 78455-78460, 78464, 78466, 78491, 78499, 78600, 78605, 78607, G0030, G0032, G0033, G0235, S9023
Other:			
Other OR therapeutic procedures on skin and breast[2]	175	85.24, 85.25, 85.31, 85.32, 85.33, 85.34, 85.35, 85.36, 85.50, 85.53, 85.54, 85.55, 85.6, 85.7, 85.70, 85.71, 85.72, 85.73, 85.74, 85.75, 85.76, 85.79, 85.82, 85.83, 85.84, 85.85, 85.86, 85.87, 85.89, 85.93, 85.94, 85.95, 85.96, 85.99, 86.06, 86.21, 86.25, 86.81, 86.82, 86.83, 86.84, 86.85, 86.86, 86.87, 86.89, 86.90, 86.91, 86.93	0061T, 11760-11762, 11770-11772, 11960-11971, 15740, 15777-15787, 15810-15819, 15824-15839, 15847, 15860-15879, 16035, 16036, 17999, 19020, 19112, 19140, 19296, 19298, 19316-19380, 19499, 20100-20103, 20926, 26560-26562, 26596, 26597, 27086, 28190, 28280, 20120-30124, 36350, 36351, 61215, 62367, 62368, S2066, S2068

Notes: [1]Index date procedure
[2]Required ICD-9 code of breast- or cancer- related diagnosis for classification (see Appendix Table 1A)
[3]Required provider to be a breast specialist